PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/51, C12Q 1/68
C12N 15/40, C12Q 1/70
A61K 39/29, C07K 13/00
G01N 33/576

A2

(11) International Publication Number: WO 92/19743

(43) International Publication Date: 12 November 1992 (12.11.92)

(21) International Application Number: PCT/US92/04036

(22) International Filing Date: 8 May 1992 (08.05.92)

(30) Priority data:
697,326 8 May 1991 (08.05.91) US

(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors: CHA, Tai-An ; 964 Springview Circle, San Ramon, CA 94583 (US). BEALL, Eileen ; 1150 Lincoln Avenue, # 5, Walnut Creek, CA 94596 (US). IRVINE, Bruce ; 3401 El Monte Drive, Concord, CA 94519 (US). KOLBERG, Janice ; 131 Scots Valley, Hercules, CA 94547 (US). URDEA, Michael, S. ; 100 Bunce Meadow Road, Alamo, CA 94501 (US).

(74) Agent: JANIUK, Anthony, J.; Wolf, Greenfield & Sacks, 600 Atlantic Avenue, Boston, MA 02210 (US).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, RU, SD, SE (European patent), SN (OAPI patent), TD (OAPI patent), TG (OAPI patent).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

(57) Abstract

The present application features nucleic acid, peptide and antibody compositions relating to genotypes of hepatitis C virus and methods of using such compositions for diagnostic and therapeutic purposes.

FILE COPY
DO NOT REMOVE

G IV = type 3
pp 24-25.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MC	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	RU	Russian Federation
CG	Congo	KP	Democratic People's Republic of Korea	SD	Sudan
CH	Switzerland	KR	Republic of Korea	SE	Sweden
CI	Côte d'Ivoire	LI	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
DE	Germany	MC	Monaco	TG	Togo
DK	Denmark			US	United States of America

HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

This application is a continuation-in-part of U.S. Serial No. 07/697,326 entitled "Polynucleotide Probes Useful for Screening for Hepatitis C Virus, filed May 8, 1991.

Technical Field

The invention relates to compositions and methods for the detection and treatment of hepatitis C virus, (HCV) infection, formerly referred to as blood-borne non-A, non-B hepatitis virus (NANBV) infection. More specifically, embodiments of the present invention feature compositions and methods for the detection of HCV, and for the development of vaccines for the prophylactic treatment of infections of HCV, and development of antibody products for conveying passive immunity to HCV.

Background of the Invention

The prototype isolate of HCV was characterized in U.S. Patent Application Serial No. 122,714 (See also EPO Publication No. 318,216). As used herein, the term "HCV" includes new isolates of the same viral species. The term "HCV-1" referred to in U.S. Patent Application Serial No. 122,714.

HCV is a transmissible disease distinguishable from other forms of viral-associated liver diseases, including that caused by the known hepatitis viruses, i.e., hepatitis A virus (HAV), hepatitis B virus (HBV), and delta hepatitis virus (HDV), as well as the hepatitis induced by cytomegalovirus (CMV) or Epstein-Barr virus (EBV). HCV was first identified in individuals who had received blood transfusions.

The demand for sensitive, specific methods for screening and identifying carriers of HCV and HCV contaminated blood or blood products is significant. Post-transfusion hepatitis (PTH) occurs in approximately 10% of transfused patients, and HCV accounts for up to 90% of these cases. The disease frequently progresses to chronic liver damage (25-55%).

Patient care as well as the prevention of transmission of HCV by blood and blood products or by close personal contact require reliable screening, diagnostic and prognostic tools to detect nucleic acids, antigens and antibodies related to HCV.

Information in this application suggests the HCV has several genotypes. That is, the genetic information of the HCV virus may not be totally identical for all HCV, but encompasses groups with differing genetic information.

Genetic information is stored in thread-like molecules of DNA and RNA. DNA consists of covalently

linked chains of deoxyribonucleotides and RNA consists of covalently linked chains of ribonucleotides. Each nucleotide is characterized by one of four bases: adenine (A), guanine (G), thymine (T), and cytosine (C). The bases are complementary in the sense that, due to the orientation of functional groups, certain base pairs attract and bond to each other through hydrogen bonding and π-stacking interactions. Adenine in one strand of DNA pairs with thymine in an opposing complementary strand. Guanine in one strand of DNA pairs with cytosine in an opposing complementary strand. In RNA, the thymine base is replaced by uracil (U) which pairs with adenine in an opposing complementary strand. The genetic code of living organism is carried in the sequence of base pairs. Living cells interpret, transcribe and translate the information of nucleic acid to make proteins and peptides.

The HCV genome is comprised of a single positive strand of RNA. The HCV genome possesses a continuous, translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the structural protein(s) appear to be encoded in approximately the first quarter of the N-terminus region, with the majority of the polyprotein responsible for non-structural proteins.

The HCV polyprotein comprises, from the amino terminus to the carboxy terminus, the nucleocapsid protein (C), the envelope protein (E), and the non-structural proteins (NS) 1, 2 (b), 3, 4 (b), and 5.

HCV of differing genotypes may encode for proteins which present an altered response to host immune systems. HCV of differing genotypes may be difficult to detect by immuno diagnostic techniques and nucleic acid probe techniques which are not specifically directed to such genotype.

Definitions for selected terms used in the application are set forth below to facilitate an understanding of the invention. The term "corresponding" means homologous to or complementary to a particular sequence of nucleic acid. As between nucleic acids and peptides, corresponding refers to amino acids of a peptide in an order derived from the sequence of a nucleic acid or its complement.

The term "non-naturally occurring nucleic acid" refers to a portion of genomic nucleic acid, cDNA, semisynthetic nucleic acid, or synthetic origin nucleic acid which, by virtue of its origin or manipulation: (1) is not associated with all of a nucleic acid with which it is associated in nature, (2) is linked to a nucleic acid or other chemical agent other than that to

which it is linked in nature, or (3) does not occur in nature.

Similarly the term, "a non-naturally occurring peptide" refers to a portion of a large naturally occurring peptide or protein, or semi-synthetic or synthetic peptide, which by virtue of its origin or manipulation (1) is not associated with all of a peptide with which it is associated in nature, (2) is linked to peptides, functional groups or chemical agents other than that to which it is linked in nature, or (3) does not occur in nature.

The term "primer" refers to a nucleic acid which is capable of initiating the synthesis of a larger nucleic acid when placed under appropriate conditions. The primer will be completely or substantially complementary to a region of the nucleic acid to be copied. Thus, under conditions conducive to hybridization, the primer will anneal to a complementary region of a larger nucleic acid. Upon addition of suitable reactants, the primer is extended by the polymerizing agent to form a copy of the larger nucleic acid.

The term "binding pair" refers to any pair of molecules which exhibit mutual affinity or binding capacity. For the purposes of the present application, the term "ligand" will refer to one molecule of the binding pair, and the term "antiligand" or "receptor"

or "target" will refer to the opposite molecule of the binding pair. For example, with respect to nucleic acids, a binding pair may comprise two complementary nucleic acids. One of the nucleic acids may be designated the ligand and the other strand is designated the antiligand receptor or target. The designation of ligand or antiligand is a matter of arbitrary convenience. Other binding pairs comprise, by way of example, antigens and antibodies, drugs and drug receptor sites and enzymes and enzyme substrates, to name a few.

The term "label" refers to a molecular moiety capable of detection including, by way of example, without limitation, radioactive isotopes, enzymes, luminescent agents, precipitating agents, and dyes.

The term "support" includes conventional supports such as filters and membranes as well as retrievable supports which can be substantially dispersed within a medium and removed or separated from the medium by immobilization, filtering, partitioning, or the like. The term "support means" refers to supports capable of being associated to nucleic acids, peptides or antibodies by binding partners, or covalent or noncovalent linkages.

A number of HCV strains and isolates have been identified. When compared with the sequence of the original isolate derived from the USA ("HCV-1"; see

Q.-L. Choo et al. (1989) Science 244:359-362, Q.-L. Choo et al. (1990) Brit. Med. Bull. 46:423-441, Q.-L. Choo et al., Proc. Natl. Acad. Sci. 88:2451-2455 (1991), and E.P.O. Patent Publication No. 318,216, cited supra), it was found that a Japanese isolate ("HCV J1") differed significantly in both nucleotide and polypeptide sequence within the NS3 and NS4 regions. This conclusion was later extended to the NS5 and envelope (E1/S and E2/NS1) regions (see K. Takeuchi et al., J. Gen. Virol. (1990) 71:3027-3033, Y. Kubo, Nucl. Acids. Res. (1989) 17:10367-10372, and K. Takeuchi et al., Gene (1990) 91:287-291). The former group of isolates, originally identified in the United States, is termed "Genotype I" throughout the present disclosure, while the latter group of isolates, initially identified in Japan, is termed "Genotype II" herein.

Brief Description of the Invention

The present invention features compositions of matter comprising nucleic acids and peptides corresponding to the HCV viral genome which define different genotypes. The present invention also features methods of using the compositions corresponding to sequences of the HCV viral genome which define different genotypes described herein.

A. Nucleic acid compositions

The nucleic acid of the present invention, corresponding to the HCV viral genome which define different genotypes, have utility as probes in nucleic acid hybridization assays, as primers for reactions involving the synthesis of nucleic acid, as binding partners for separating HCV viral nucleic acid from other constituents which may be present, and as anti-sense nucleic acid for preventing the transcription or translation of viral nucleic acid.

One embodiment of the present invention features a composition comprising a non-naturally occurring nucleic acid having a nucleic acid sequence of at least eight nucleotides corresponding to a non-HCV-1 nucleotide sequence of the hepatitis C viral genome. Preferably, the nucleotide sequence is selected from a sequence present in at least one region consisting of the NS5 region, envelope 1 region, 5'UT region, and the core region.

Preferably, with respect to sequences which correspond to the NS5 region, the sequence is selected from a sequence within a sequence numbered 2-22. The sequence numbered 1 corresponds to HCV-1. Sequences numbered 1-22 are defined in the Sequence Listing of the application.

Preferably, with respect to sequences corresponding to the envelope 1 region, the sequence is

selected from a sequence within sequences numbered 24-32. Sequence No. 23 corresponds to HCV-1. Sequences numbered 23-32 are set forth in the Sequence Listing of the application.

Preferably, with respect to the sequences which correspond to the 5'UT regions, the sequence is selected from a sequence within sequences numbered 34-51. Sequence No. 33 corresponds to HCV-1. Sequence No. 33-51 are set forth in the Sequence Listing of this application.

Preferably, with respect to the sequences which correspond to the core region, the sequence is selected from a sequence within the sequences numbered 53-66. Sequence No. 52 corresponds to HCV-1. Sequences 52-66 are set forth in the Sequence Listing of this application.

The compositions of the present invention form hybridization products with nucleic acid corresponding to different genotypes of HCV.

HCV has at least five genotypes, which will be referred to in this application by the designations GI-GV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV,

is exemplified by sequences numbered 20-22, and 29-31 and 48-49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

One embodiment of the present invention features compositions comprising a nucleic acid having a sequence corresponding to one or more sequences which exemplify a genotype of HCV.

B. Method of forming a Hybridization Product

Embodiments of the present invention also feature a method of forming a hybridization product with nucleic acid having a sequence corresponding to HCV nucleic acid. One method comprises the steps of placing a non-naturally occurring nucleic acid having a non-HCV-1 sequence corresponding to HCV nucleic acid under conditions in which hybridization may occur. The non-naturally occurring nucleic acid is capable of forming a hybridization product with HCV nucleic acid, under hybridization conditions. The method further comprises the step of imposing hybridization conditions to form a hybridization product in the presence of nucleic acid corresponding to a region of the HCV genome.

The formation of a hybridization product has utility for detecting the presence of one or more genotypes of HCV. Preferably, the non-naturally occurring nucleic acid forms a hybridization product

with nucleic acid of HCV in one or more regions comprising the NS5 region, envelope 1 region, 5'UT region and the core region. To detect the hybridization product, it is useful to associate the non-naturally occurring nucleic acid with a label. The formation of the hybridization product is detected by separating the hybridization product from labeled non-naturally occurring nucleic acid, which has not formed a hybridization product.

The formation of a hybridization product has utility as a means of separating one or more genotypes of HCV nucleic acid from other constituents potentially present. For such applications, it is useful to associate the non-naturally occurring nucleic acid with a support for separating the resultant hybridization product from the the other constituents.

Nucleic acid "sandwich assays" employ one nucleic acid associated with a label and a second nucleic acid associated with a support. An embodiment of the present invention features a sandwich assay comprising two nucleic acids, both have sequences which correspond to HCV nucleic acids; however, at least one non-naturally occurring nucleic acid has a sequence corresponding to non-HCV-1 HCV nucleic acid. At least one nucleic acid is capable of associating with a label, and the other is capable of associating with a support. The support associated non-naturally

occurring nucleic acid is used to separate the hybridization products which include an HCV nucleic acid and the non-naturally occurring nucleic acid having a non-HCV-1 sequence.

One embodiment of the present invention features a method of detecting one or more genotypes of HCV. The method comprises the steps of placing a non-naturally occurring nucleic acid under conditions which hybridization may occur. The non-naturally occurring nucleic acid is capable of forming a hybridization product with nucleic acid from one or more genotypes of HCV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified sequences numbered 20-22 and 29-31. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

The hybridization product of HCV nucleic acid with a non-naturally occurring nucleic acid having non-HCV-1 sequence corresponding to sequences within the HCV genome has utility for priming a reaction for the synthesis of nucleic acid.

The hybridization product of HCV nucleic acid with a non-naturally occurring nucleic acid having a

sequence corresponding to a particular genotype of HCV has utility for priming a reaction for the synthesis of nucleic acid of such genotype. In one embodiment, the synthesized nucleic acid is indicative of the presence of one or more genotypes of HCV.

The synthesis of nucleic acid may also facilitate cloning of the nucleic acid into expression vectors which synthesize viral proteins.

Embodiments of the present methods have utility as anti-sense agents for preventing the transcription or translation of viral nucleic acid. The formation of a hybridization product of a non-naturally occurring nucleic acid having sequences which correspond to a particular genotype of HCV genomic sequencing with HCV nucleic acid may block translation or transcription of such genotype. Therapeutic agents can be engineered to include all five genotypes for inclusivity.

C. Peptide and antibody composition

A further embodiment of the present invention features a composition of matter comprising a non-naturally occurring peptide of three or more amino acids corresponding to a nucleic acid having a non-HCV-1 sequence. Preferably, the non-HCV-1 sequence corresponds with a sequence within one or more regions consisting of the NS5 region, the envelope 1 region, the 5'UT region, and the core region.

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence of the NS5 region, the sequence is within sequences numbered 2-22. The sequence numbered 1 corresponds to HCV-1. Sequences numbered 1-22 are set forth in the Sequence Listing.

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence of the envelope 1 region, the sequence is within sequences numbered 24-32. The sequence numbered 23 corresponds to HCV-1. Sequences numbered 23-32 are set forth in the Sequence Listing.

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence directed to the core region, the sequence is within sequences numbered 53-66. Sequence numbered 52 corresponds to HCV-1. Sequences numbered 52-66 are set forth in the Sequence Listing.

The further embodiment of the present invention features peptide compositions corresponding to nucleic acid sequences of a genotype of HCV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified

sequences numbered 20-22, 29-31, 48 and 49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

The non-naturally occurring peptides of the present invention are useful as a component of a vaccine. The sequence information of the present invention permits the design of vaccines which are inclusive for all or some of the different genotypes of HCV. Directing a vaccine to a particular genotype allows prophylactic treatment to be tailored to maximize the protection to those agents likely to be encountered. Directing a vaccine to more than one genotype allows the vaccine to be more inclusive.

The peptide compositions are also useful for the development of specific antibodies to the HCV proteins. One embodiment of the present invention features as a composition of matter, an antibody to peptides corresponding to a non-HCV-1 sequence of the HCV genome. Preferably, the non-HCV-1 sequence is selected from the sequence within a region consisting of the NS5 region, the envelope 1 region, and the core region. There are no peptides associated with the untranslated 5'UT region.

Preferably, with respect to antibodies directed to peptides of the NS5 region, the peptide corresponds to a sequence within sequences numbered 2-22. Preferably, with respect to antibodies directed to a peptide

corresponding to the envelope 1 region, the peptide corresponds to a sequence within sequences numbered 24-32. Preferably, with respect to the antibodies directed to peptides corresponding to the core region, the peptide corresponds to a sequence within sequences numbered 53-66.

Antibodies directed to peptides which reflect a particular genotype have utility for the detection of such genotypes of HCV and therapeutic agents..

One embodiment of the present invention features an antibody directed to a peptide corresponding to nucleic acid having sequences of a particular genotype. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified sequences numbered 20-22, 29-31, 48 and 49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

Individuals skilled in the art will readily recognize that the compositions of the present invention can be packaged with instructions for use in the form of a kit for performing nucleic acid hybridizations or immunochemical reactions.

The present invention is further described in the following figures which illustrate sequences demonstrating genotypes of HCV. The sequences are designated by numerals 1-145, which numerals and sequences are consistent with the numerals and sequences set forth in the Sequence Listing. Sequences 146 and 147 facilitate the discussion of an assay which numerals and sequences are consistent with the numerals and sequences set forth in the Sequence Listing.

Brief Description of the Figures and Sequence Listing

Figure 1 depicts schematically the genetic organization of HCV;

Figure 2 sets forth nucleic acid sequences numbered 1-22 which sequences are derived from the NS5 region of the HCV viral genome;

Figure 3 sets forth nucleic acid sequences numbered 23-32 which sequences are derived from the envelope 1 region of the HCV viral genome;

Figure 4 sets forth nucleic acid sequences numbered 33-51 which sequences are derived from the 5'UT region of the HCV viral genome; and,

Figure 5 sets forth nucleic acid sequences numbered 52-66 which sequences are derived from the core region of the HCV viral genome.

The Sequence Listing sets forth the sequences of sequences numbered 1-147.

Detailed Description of the Invention

The present invention will be described in detail as as nucleic acid having sequences corresponding to the HCV genome and related peptides and binding partners, for diagnostic and therapeutic applications.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitsch & Sambrook, Molecular Cloning; A Laboratory Manual (1982); DNA Cloning, Volumes I and II (D.N Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed, 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); the series, Methods in Enzymology (Academic Press, Inc.), particularly Vol. 154 and Vol. 155 (Wu and Grossman, eds.).

The cDNA libraries are derived from nucleic acid sequences present in the plasma of an HCV-infected chimpanzee. The construction of one of these libraries, the "c" library (ATCC No. 40394), is described in PCT Pub. No. WO90/14436. The sequences of the library relevant to the present invention are set forth herein as sequence numbers 1, 23, 33 and 52.

Nucleic acids isolated or synthesized in accordance with features of the present invention are

useful, by way of example without limitation as probes, primers, anti-sense genes and for developing expression systems for the synthesis of peptides corresponding to such sequences.

The nucleic acid sequences described define genotypes of HCV with respect to four regions of the viral genome. Figure 1 depicts schematically the organization of HCV. The four regions of particular interest are the NS5 region, the envelope 1 region, the 5'UT region and the core region.

The sequences set forth in the present application as sequences numbered 1-22 suggest at least five genotypes in the NS5 region. Sequences numbered 1-22 are depicted in Figure 2 as well as the Sequence Listing. Each sequence numbered 1-22 is derived from nucleic acid having 340 nucleotides from the NS5 region.

The five genotypes are defined by groupings of the sequences defined by sequence numbered 1-22. For convenience, in the present application, the different genotypes will be assigned roman numerals and the letter "G".

The first genotype (GI) is exemplified by sequences within sequences numbered 1-6. A second genotype (GII) is exemplified by sequences within sequences numbered 7-12. A third genotype (GIII) is exemplified by the sequences within sequences numbered 13-17. A fourth genotype (GIV) is exemplified by

sequences within sequences numbered 20-22. A fifth genotype (GV) is exemplified by sequences within sequences numbered 18 and 19.

The sequences set forth in the present application as sequences numbered 23-32 suggest at least four genotypes in the envelope 1 region of HCV. Sequences numbered 23-32 are depicted in Figure 3 as well as in the Sequence Listing. Each sequence numbered 23-32 is derived from nucleic acid having 100 nucleotides from the envelope 1 region.

A first envelope 1 genotype group (GI) is exemplified by the sequences within the sequences numbered 23-25. A second envelope 1 genotype (GII) region is exemplified by sequences within sequences numbered 26-28. A third envelope 1 genotype (GIII) is exemplified by the sequences within sequences numbered 32. A fourth envelope 1 genotype (GIV) is exemplified by the sequences within sequence numbered 29-31.

The sequences set forth in the present application as sequences numbered 33-51 suggest at least three genotypes in the 5'UT region of HCV. Sequences numbered 33-51 are depicted in Figure 4 as well as in the Sequence Listing. Each sequence numbered 33-51 is derived from the nucleic acid having 252 nucleotides from the 5'UT region, although sequences 50 and 51 are somewhat shorter at approximately 180 nucleotides.

The first 5'UT genotype (GI) is exemplified by the sequences within sequences numbered 33-38. A second 5'UT genotype (GII) is exemplified by the sequences within sequences numbered 39-45. A third 5'UT genotype (GIII) is exemplified by the sequences within sequences numbered 46-47. A fourth 5'UT genotype (GIV) is exemplified by sequences within sequences humbered 48 and 49. A fifth 5'UT genotype (GV) is exemplified by sequences within sequences numbered 50 and 51.

The sequences numbered 48-62 suggest at least three genotypes in the core region of HCV. The sequences numbered 52-66 are depicted in Figure 5 as well as in the Sequence Listing.

The first core region genotype (GI) is exemplified by the sequences within sequences numbered 52-57. The second core region genotype (GII) is exemplified by sequences within sequences numbered 58-64. The third core region genotype (GIII) is exemplified by sequences within sequences numbered 65 and 66. Sequences numbered 52-65 are comprised of 549 nucleotides. Sequence numbered 66 is comprised of 510 nucleotides.

The various genotypes described with respect to each region are consistent. That is, HCV having features of the first genotype with respect to the NS5 region will substantially conform to features of the first genotype of the envelope 1 region, the 5'UT region and the core region.

Nucleic acid isolated or synthesized in accordance with the sequences set forth in sequence numbers 1-66 are useful as probes, primers, capture ligands and anti-sense agents. As probes, primers, capture ligands and anti-sense agents, the nucleic acid wil normally comprise approximately eight or more nucleotides for specificity as well as the ability to form stable hybridization products.

Probes

A nucleic acid isolated or synthesized in accordance with a sequence defining a particular genotype of a region of the HCV genome can be used as a probe to detect such genotype or used in combination with other nucleic acid probes to detect substantially all genotypes of HCV.

With the sequence information set forth in the present application, sequences of eight or more nucleotides are identified which provide the desired inclusivity and exclusivity with respect to various genotypes within HCV, and extraneous nucleic acid sequences likely to be encountered during hybridization conditions.

Individuals skilled in the art will readily recognize that the nucleic acid sequences, for use as probes, can be provided with a label to facilitate detection of a hybridization product.

Capture Ligand

For use as a capture ligand, the nucleic acid selected in the manner described above with respect to probes, can be readily associated with supports. The manner in which nucleic acid is associated with supports is well known. Nucleic acid having sequences corresponding to a sequence within sequences numbered 1-66 have utility to separate viral nucleic acid of one genotype from the nucleic acid of HCV of a different genotype. Nucleic acid isolated or synthesized in accordance with sequences within sequences numbered 1-66, used in combinations, have utility to capture substantially all nucleic acid of all HCV genotypes.

Primers

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as primers for the amplification of HCV sequences. With respect to polymerase chain reaction (PCR) techniques, nucleic acid sequences of eight or more nucleotides corresponding to one or more sequences of sequences numbered 1-66 have utility in conjunction with suitable enzymes and reagents to create copies of the viral nucleic acid. A plurality of primers having different sequences corresponding to more than one genotype can be used to create copies of viral nucleic acid for such genotypes.

The copies can be used in diagnostic assays to detect HCV virus. The copies can also be incorporated into cloning and expression vectors to generate polypeptides corresponding to the nucleic acid synthesized by PCR, as will be described in greater detail below.

Anti-sense

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as anti-sense genes to prevent the expression of HCV.

Nucleic acid corresponding to a genotype of HCV is loaded into a suitable carrier such as a liposome for introduction into a cell infected with HCV. A nucleic acid having eight or more nucleotides is capable of binding to viral nucleic acid or viral messenger RNA. Preferably, the anti-sense nucleic acid is comprised of 30 or more nucleotides to provide necessary stability of a hybridization product of viral nucleic acid or viral messenger RNA. Methods for loading anti-sense nucleic acid is known in the art as exemplified by U.S. Patent 4,241,046 issued December 23, 1980 to Papahadjopoulos et al.

Peptide Synthesis

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility to

generate peptides. The sequences exemplified by sequences numbered 1-32 and 52-66 can be cloned into suitable vectors or used to isolate nucleic acid. The isolated nucleic acid is combined with suitable DNA linkers and cloned into a suitable vector. The vector can be used to transform a suitable host organism such as E. coli and the peptide encoded by the sequences isolated.

Molecular cloning techniques are described in the text Molecular Cloning: A Laboratory Manual, Maniatis et al., Coldspring Harbor Laboratory (1982).

The isolated peptide has utility as an antigenic substance for the development of vaccines and antibodies directed to the particular genotype of HCV.

Vaccines and Antibodies

The peptide materials of the present invention have utility for the development of antibodies and vaccines.

The availability of cDNA sequences, or nucleotide sequences derived therefrom (including segments and modifications of the sequence), permits the construction of expression vectors encoding antigenically active regions of the peptide encoded in either strand. The antigenically active regions may be derived from the NS5 region, envelope 1 regions, and the core region.

Fragments encoding the desired peptides are derived from the cDNA clones using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and vectors which are useful for the production of polypeptides which contain fusion sequences of SOD are described in European Patent Office Publication number 0196056, published October 1, 1986.

Any desired portion of the HCV cDNA containing an open reading frame, in either sense strand, can be obtained as a recombinant peptide, such as a mature or fusion protein; alternatively, a peptide encoded in the cDNA can be provided by chemical synthesis.

The DNA encoding the desired peptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, may be ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host systems are presently used in forming recombinant peptides. The peptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, differential extraction, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and

the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins. Such peptides can be used as diagnostics, or those which give rise to neutralizing antibodies may be formulated into vaccines. Antibodies raised against these peptides can also be used as diagnostics, or for passive immunotherapy or for isolating and identifying HCV.

An antigenic region of a peptide is generally relatively small--typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to NS5 region, envelope 1 region, and the core region of the HCV genome. The 5'UT region is not known to be translated. Accordingly, using the cDNAs of such regions, DNAs encoding short segments of HCV peptides corresponding to such regions can be expressed recombinantly either as fusion proteins, or as isolated peptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis. In instances wherein the synthesized peptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the peptide may be linked to a suitable carrier.

A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-

pyridylthio)propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue). These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example, Immun Rev (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaprioc acid, 2-bromoacetic acid, 2-iodoacetic acid, 4-N-maleimido-methyl)cyclohexane-1-carboxylic acid, and the like. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2 nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

Any carrier may be used which does not itself induce the production of antibodies harmful to the

host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized Sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

Peptides comprising HCV amino acid sequences encoding at least one viral epitope derived from the NS5, envelope 1, and core region are useful immunological reagents. The 5'UT region is not known to be translated. For example, peptides comprising such truncated sequences can be used as reagents in an immunoassay. These peptides also are candidate subunit antigens in compositions for antiserum production or vaccines. While the truncated sequences can be produced by various known treatments of native viral protein, it is generally preferred to make synthetic or recombinant peptides comprising HCV sequence. Peptides comprising these truncated HCV sequences can be made up entirely of HCV sequences (one or more epitopes, either contiguous or noncontiguous), or HCV sequences and heterologous sequences in a fusion protein. Useful

heterologous sequences include sequences that provide for secretion from a recombinant host, enhance the immunological reactivity of the HCV epitope(s), or facilitate the coupling of the polypeptide to an immunoassay support or a vaccine carrier. See, E.G., EPO Pub. No. 116,201; U.S. Pat. No. 4,722,840; EPO Pub. No. 259,149; U.S. Pat. No. 4,629,783.

The size of peptides comprising the truncated HCV sequences can vary widely, the minimum size being a sequence of sufficient size to provide an HCV epitope, while the maximum size is not critical. For convenience, the maximum size usually is not substantially greater than that required to provide the desired HCV epitopes and function(s) of the heterologous sequence, if any. Typically, the truncated HCV amino acid sequence will range from about 5 to about 100 amino acids in length. More typically, however, the HCV sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select HCV sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids.

HCV amino acid sequences comprising epitopes can be identified in a number of ways. For example, the entire protein sequence corresponding to each of the NS5, envelope 1, and core regions can be screened by preparing a series of short peptides that together span

the entire protein sequence of such regions. By starting with, for example, peptides of approximately 100 amino acids, it would be routine to test each peptide for the presence of epitope(s) showing a desired reactivity, and then testing progressively smaller and overlapping fragments from an identified peptides of 100 amino acids to map the epitope of interest. Screening such peptides in an immunoassay is within the skill of the art. It is also known to carry out a computer analysis of a protein sequence to identify potential epitopes, and then prepare peptides comprising the identified regions for screening.

The immunogenicity of the epitopes of HCV may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B surface antigen. See, e.g., US 4,722,840. Constructs wherein the HCV epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the HCV epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles constructed from particle forming protein which include HCV sequences are immunogenic with respect to HCV and HBV.

Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in S. cerevisiae (P. Valenzuela et al. (1982)), as well as in, for example, mammalian cells (P. Valenzuela et al. 1984)). The formation of such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath et al. (1984). Constructs of the pre-S-HBSAg particle expressible in yeast are disclosed in EPO 174,444, published March 19, 1986; hybrids including heterologous viral sequences for yeast expression are disclosed in EPO 175,261, published March 26, 1966. These constructs may also be expressed in mammalian cells such as Chinese hamster ovary (CHO) cells using an SV40-dihydrofolate reductase vector (Michelle et al. (1984)).

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an HCV epitope. In this replacement, regions which are not required to mediate the aggregation of the units to form immunogenic particles in yeast of mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the HCV epitope.

Vaccines

Vaccines may be prepared from one or more

immunogenic peptides derived from HCV. The observed homology between HCV and Flaviviruses provides information concerning the peptides which are likely to be most effective as vaccines, as well as the regions of the genome in which they are encoded.

Multivalent vaccines against HCV may be comprised of one or more epitopes from one or more proteins derived from the NS5, envelope 1, and core regions. In particular, vaccines are contemplated comprising one or more HCV proteins or subunit antigens derived from the NS5, envelope 1, and core regions. The 5'UT region is not known to be translated.

The preparation of vaccines which contain an immunogenic peptide as an active ingredient, is known to one skilled in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or

emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-theronyl-D- isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl- D-isoglutamine (CGP 11637, referred to as nor-MDP), N- acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1- 2-dipalmitoyl -sn-glycero-3-hydroxyphosphoryloxy)- ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic peptide containing an HCV antigenic sequence resulting from administration of this peptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such

suppositories may be formed from mixtures containing the active ingredient in the range of 0/5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

The examples below are provided for illustrative purposes and are not intended to limit the scope of the present invention.

I. Detection of HCV RNA from Serum

RNA was extracted from serum using guanidinium salt, phenol and chloroform according to the instructions of the kit manufacturer (RNAzol™ B kit, Cinna/Biotecx). Extracted RNA was precipitated with isopropanol and washed with ethanol. A total of 25 µl serum was processed for RNA isolation, and the purified RNA was resuspended in 5 µl diethyl pyrocarbonate treated water for subsequent cDNA synthesis.

II. cDNA Synthesis and Polymerase Chain Reaction (PCR) Amplification

Table 1 lists the sequence and position (with reference to HCV1) of all the PCR primers and probes used in these examples. Letter designations for

nucleotides are consistent with 37 C.F.R. §§1.821-1.825. Thus, the letters A, C, G, T, and U are used in the ordinary sense of adenine, cytosine, guanine, thymine, and uracil. The letter M means A or C; R means A or G; W means A or T/U; S means C or G; Y means C or T/U; K means G or T/U; V means A or C or G, not T/U; H means A or C or T/U, not G; D means A or G or T/U, not C; B means C or G or T/U, not A; N means (A or C or G or T/U) or (unknown or other). Table 1 is set forth below:

Table 1

Seq. No.	Sequence (5'-3')	Nucleotide Position
67	CAAACGTAACACCAACCGRCGCCCACAGG	374-402
68	ACAGAYCCGCAKAGRTCCCCCACG	1192-1169
69	GCAACCTCGAGGTAGACGTCAGCCTATCCC	509-538
70	GCAACCTCGTGGAAGGCGACAACCTATCCC	509-538
71	GTCACCAATGATTGCCCTAACTCGAGTATT	948-977
72	GTCACGAACGACTGCTCCAACTCAAG	948-973
73	TGGACATGATCGCTGGWGCYCACTGGGG	1375-1402
74	TGGAYATGGTGGYGGGGGCYCACTGGGG	1375-1402
75	ATGATGAACTGGTCVCCYAC	1308-1327
76	ACCTTVGCCCAGTTSCCCRCCATGGA	1453-1428
77	AACCCACTCTATGYCCGGYCAT	205-226
78	GAATCGCTGGGGTGACCG	171-188
79	CCATGAATCACTCCCCTGTGAGGAACTA	30-57
80	TTGCGGGGGCACGCCCAA	244-227

For cDNA synthesis and PCR amplification, a protocol developed by Perkin-Elmer/Cetus (GeneAmp® RNA PCR kit) was used. Both random hexamer and primers with specific complementary sequences to HCV were employed to prime the reverse transcription (RT) reaction. All processes, except for adding and mixing reaction components, were performed in a thermal cycler (MJ Research, Inc.). The first strand cDNA synthesis reaction was inactivated at 99°C for 5 min, and then cooled at 50°C for 5 min before adding reaction components for subsequent amplification. After an initial 5 cycles of 97°C for 1 min, 50°C for 2 min, and 72°C for 3 min, 30 cycles of 94°C for 1 min, 55°C for 2 min, and 72°C for 3 min followed, and then a final 7 min of elongation at 72°C.

For the genotyping analysis, sequences 67 and 68 were used as primers in the PCR reaction. These primers amplify a segment corresponding to the core and envelope regions. After amplification, the reaction products were separated on an agarose gel and then transferred to a nylon membrane. The immobilized reaction products were allowed to hybridize with a ^{32}P-labelled nucleic acid corresponding to either Genotype I (core or envelope 1) or Genotype II (core or envelope 1). Nucleic acid corresponding to Genotype 1 comprised sequences numbered 69 (core), 71 (envelope), and 73 (envelope). Nucleic acid corresponding to

Genotype II comprised sequences numbered 70 (core), 72 (envelope), and 74 (envelope).

The Genotype I probes only hybridized to the product amplified from isolates which had Genotype I sequence. Similarly, Genotype II probes only hybridized to the product amplified from isolates which had Genotype II sequence.

In another experiment, PCR products were generated using sequences 79 and 80. The products were analyzed as described above except Sequence No. 73 was used to detect Genotype I, Sequence No. 74 was used to detect Genotype II, Sequence No. 77 (5'UT) was used to detect Genotype III, and Sequence No. 78 (5'UT) was used to detect Genotype IV. Each sequence hybridized in a genotype specific manner.

III. Detection of HCV GI-GIV using a sandwich hybridization assay for HCV RNA

An amplified solution phase nucleic acid sandwich hybridization assay format is described in this example. The assay format employs several nucleic acid probes to effect capture and detection. A capture probe nucleic acid is capable of associating a complementary probe bound to a solid support and HCV nucleic acid to effect capture. A detection probe nucleic acid has a first segment (A) that binds to HCV nucleic acid and a second segment (B) that hybridizes to a second amplifier nucleic acid.

The amplifier nucleic acid has a first segment (B*) that hybridizes to segment (B) of the probe nucleic acid and also comprises fifteen iterations of a segment (C). Segment C of the amplifier nucleic acid is capable of hybridizing to three labeled nucleic acids.

Nucleic acid sequences which correspond to nucleotide sequences of the envelope 1 gene of Group I HCV isolates are set forth in sequences numbered 81-99. Table 2 sets forth the area of the HCV genome to which the nucleic acid sequences correspond and a preferred use of the sequences.

Table 2

Probe Type	Sequence No.	Complement of Nucleotide Numbers
Label	81	879-911
Label	82	912-944
Capture	83	945-977
Label	84	978-1010
Label	85	1011-1043
Label	86	1044-1076
Label	87	1077-1109
Capture	88	1110-1142
Label	89	1143-1175

Table 2 continued

Probe Type	Sequence No.	Complement of Nucleotide Numbers
Label	90	1176-1208
Label	91	1209-1241
Label	92	1242-1274
Capture	93	1275-1307
Label	94	1308-1340
Label	95	1341-1373
Label	96	1374-1406
Label	97	1407-1439
Capture	98	1440-1472
Label	99	1473-1505

Nucleic acid sequences which correspond to nucleotide sequences of the envelope 1 gene of Group II HCV isolates are set forth in sequences 100-118. Table 3 sets forth the area of the HCV genome to which the nucleic acid corresponds and the preferred use of the sequences.

Table 3

Probe Type	Sequence No.	Complement of Nucleotide Numbers
Label	100	879-911
Label	101	912-944
Capture	102	945-977
Label	103	978-1010
Label	104	1011-1043
Label	105	1044-1076
Label	106	1077-1109
Capture	107	1110-1142
Label	108	1143-1175
Label	109	1176-1208
Label	110	1209-1241
Label	111	1242=1274
Capture	112	1275-1307
Label	113	1308-1340
Label	114	1341-1373
Label	115	1374-1406
Label	116	1407-1439
Capture	117	1440-1472
Label	118	1473-1505

Nucleic acid sequences which correspond to nucleotide sequences in the C gene and the 5'UT region

are set forth in sequences 119-145. Table 4 identifies the sequence with a preferred use.

Table 4

Probe Type	Sequence No.
Capture	119
Label	120
Label	121
Label	122
Capture	123
Label	124
Label	125
Label	126
Capture	127
Label	128
Label	129
Label	130
Capture	131
Label	132
Label	133
Label	134
Label	135
Capture	136
Label	137
Label	138

Table 4 continued

Probe Type	Sequence No.
Label	139
Capture	140
Label	141
Label	142
Label	143
Capture	144
Label	145

The detection and capture probe HCV-specific segments, and their respective names as used in this assay were as follows.

Capture sequences are sequences numbered 119-122 and 141-144.

Detection sequences are sequences numbered 119-140.

Each detection sequence contained, in addition to the sequences substantially complementary to the HCV sequences, a 5' extension (B) which extension (B) is complementary to a segment of the second amplifier nucleic acid. The extension (B) sequence is identified in the Sequence Listing as Sequence No. 146, and is reproduced below.

AGGCATAGGACCCGTGTCTT

Each capture sequence contained, in addition to the sequences substantially complementary to HCV sequences, a sequence complementary to DNA bound to a solid phase. The sequence complementary to DNA bound to a solid support was carried downstream from the capture sequence. The sequence complementary to the DNA bound to the support is set forth as Sequence No. 147 and is reproduced below.

CTTCTTTGGAGAAAGTGGTG

Microtiter plates were prepared as follows. White Microlite 1 Removawell strips (polystyrene microtiter plates, 96 wells/plate) were purchased from Dynatech Inc.

Each well was filled with 200 µl 1 N HCl and incubated at room temperature for 15-20 min. The plates were then washed 4 times with 1X PBS and the wells aspirated to remove liquid. The wells were then filled with 200 µl 1 N NaOH and incubated at room temperature for 15-20 min. The plates were again washed 4 times with 1X PBS and the wells aspirated to remove liquid.

Poly(phe-lys) was purchased from Sigma Chemicals, Inc. This polypeptide has a 1:1 molar ratio of phe:lys and an average m.w. of 47,900 gm/mole. It has an average length of 309 amino acids and contains 155 amines/mole. A 1 mg/ml solution of the polypeptide was mixed with 2M NaCl/1X PBS to a final concentration of

0.1 mg/ml (pH 6.0). A volume of 200 µl of this solution was added to each well. The plate was wrapped in plastic to prevent drying and incubated at 30°C overnight. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid.

The following procedure was used to couple the nucleic acid, a complementary sequence to Sequence No. 147, to the plates, hereinafter referred to as immobilized nucleic acid. Synthesis of immobilized nucleic acid having a sequence complementary to Sequence No. 133 was described in EPA 883096976. A quantity of 20 mg disuccinimidyl suberate was dissolved in 300 µl dimethyl formamide (DMF). A quantity of 26 OD_{260} units of immobilized nucleic acid was added to 100 µl coupling buffer (50 mM sodium phosphate, pH 7.8). The coupling mixture was then added to the DSS-DMF solution and stirred with a magnetic stirrer for 30 min. An NAP-25 column was equilibrated with 10 mM sodium phosphate, pH 6.5. The coupling mixture DSS-DMF solution was added to 2 ml 10 mM sodium phosphate, pH 6.5, at 4°C. The mixture was vortexed to mix and loaded onto the equilibrated NAP-25 column. DSS-activated immobilized nucleic acid DNA was eluted from the column with 3.5 ml 10 mM sodium phosphate, pH 6.5. A quantity of 5.6 OD_{260} units of eluted DSS-activated immobilized nucleic acid DNA was added to 1500 ml 50 mM sodium phosphate, pH 7.8. A volume of 50

µl of this solution was added to each well and the plates were incubated overnight. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid.

Final stripping of plates was accomplished as follows. A volume of 200 µl of 0.2N NaOH containing 0.5% (w/v) SDS was added to each well. The plate was wrapped in plastic and incubated at 65°C for 60 min. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid. The stripped plate was stored with desiccant beads at 2-8°C.

Serum samples to be assayed were analyzed using PCR followed by sequence analysis to determine the genotype.

Sample preparation consisted of delivering 50 µl of the serum sample and 150 µl P-K Buffer (2 mg/ml proteinase K in 53 mM Tris-HCl, pH 8.0/0.6 M NaCl/0.06 M sodium citrate/8 mM EDTA, pH 8.0/1.3%SDS/16µg/ml sonicated salmon sperm DNA/7% formamide/50 fmoles capture probes/160 fmoles detection probes) to each well. Plates were agitated to mix the contents in the well, covered and incubated for 16 hr at 62°C.

After a further 10 minute period at room temperature, the contents of each well were aspirated to remove all fluid, and the wells washed 2X with washing buffer (0.1% SDS/0.015 M NaCl/ 0.0015 M sodium citrate). The amplifier nucleic acid was then added to

each well (50 µl of 0.7 fmole/µl solution in 0..48 M NaCl/0.048 M sodium citrate/0.1% SDS/0.5% "blocking reagent" (Boehringer Mannheim, catalog No. 1096 176)). After covering the plates and agitating to mix the contents in the wells, the plates were incubated for 30 min. at 52°C.

After a further 10 min period at room temperature, the wells were washed as described above.

Alkaline phosphatase label nucleic acid, disclosed in EP 883096976, was then added to each well (50 µl/well of 2.66 fmoles/µl). After incubation at 52°C for 15 min., and 10 min. at room temperature, the wells were washed twice as above and then 3X with 0.015 M NaCl/0.0015 M sodium citrate.

An enzyme-triggered dioxetane (Schaap et al., Tet. Lett. (1987) 28:1159-1162 and EPA Pub. No. 0254051), obtained from Lumigen, Inc., was employed. A quantity of 50 µl Lumiphos 530 (Lumigen) was added to each well. The wells were tapped lightly so that the reagent would fall to the bottom and gently swirled to distribute the reagent evenly over the bottom. The wells were covered and incubated at 37°C for 20-40 min.

Plates were then read on a Dynatech ML 1000 luminometer. Output was given as the full integral of the light produced during the reaction.

The assay positively detected each of the serum samples, regardless of genotype.

IV. Expression of the Polypeptide Encoded in Sequences Defined by Differing Genotypes

HCV polypeptides encoded by a sequence within sequences 1-66 are expressed as a fusion polypeptide with superoxide dismutase (SOD). A cDNA carrying such sequences is subcloned into the expression vector pSODcf1 (Steimer et al. 1986)).

First, DNA isolated from pSODcf1 is treated with BamHI and EcoRI, and the following linker was ligated into the linear DNA created by the restriction enzymes:

5 GAT CCT GGA ATT CTG ATA AGA
CCT TAA GAC TAT TTT AA 3

After cloning, the plasmid containing the insert is isolated.

Plasmid containing the insert is restricted with EcoRI. The HCV cDNA is ligated into this EcoRI linearized plasmid DNA. The DNA mixture is used to transform E. coli strain D1210 (Sadler et al. (1980)). Polypeptides are isolated on gels.

V. Antigenicity of Polypeptides

The antigenicity of polypeptides formed in Section IV is evaluated in the following manner. Polyethylene pins arranged on a block in an 8 12 array (Coselco Mimetopes, Victoria, Australia) are prepared by placing the pins in a bath (20% v/v piperidine in dimethylformamide (DMF)) for 30 minutes at room

temperature. The pins are removed, washed in DMF for 5 minutes, then washed in methanol four times (2 min/wash). The pins are allowed to air dry for at least 10 minutes, then washed a final time in DMF (5Min). 1-Hydroxybenzotriazole (HOBt, 367 mg) is dissolved in DMF (80 µL) for use in coupling Fmoc-protected polypeptides prepared in Section IV.

The protected amino acids are placed in micro-titer plate wells with HOBt, and the pin block placed over the plate, immersing the pins in the wells. The assembly is then sealed in a plastic bag and allowed to react at 25°C for 18 hours to couple the first amino acids to the pins. The block is then removed, and the pins washed with DMF (2 min.), MeOH (4 x, 2 min.), and again with DMF (2 min.) to clean and deprotect the bound amino acids. The procedure is repeated for each additional amino acid coupled, until all octamers are prepared.

The free N-termini are then acetylated to compensate for the free amide, as most of the epitopes are not found at the N-terminus and thus would not have the associated positive charge. Acetylation is accomplished by filling the wells of a microtiter plate with DMF/acetic anhydride/triethylamine (5:2:1 v/v/v) and allowing the pins to react in the wells for 90 minutes at 20°C. The pins are then washed with DMF (2

min.) and MeOH (4 x, 2 min.), and air dried for at least 10 minutes.

The side chain protecting groups are removed by treating the pins with trifluoroacetic acid/phenol/ dithioethane (95:2.5:1.5, v/v/v) in polypropylene bags for 4 hours at room temperature. The pins are then washed in dichloromethane (2 x, 2 min.), 5% di-isopropylethylamine/dichloromethane (2 x, 5 min.), dichloromethane (5 min.), and air-dried for at least 10 minutes. The pins are then washed in water (2 min.), MeOH (18 hours), dried in vacuo, and stored in sealed plastic bags over silica gel. IV.B.15.b Assay of Peptides.

Octamer-bearing pins are treated by sonicating for 30 minutes in a disruption buffer (1% sodium dodecylsulfate, 0.1% 2-mercaptoethanol, 0.1 M NaH_2PO_4) at 60°C. The pins are then immersed several times in water (60°C), followed by boiling MeOH (2 min.), and allowed to air dry.

The pins are then precoated for 1 hour at 25°C in microtiter wells containing 200 µL blocking buffer (1% ovalbumin, 1% BSA, 0.1% Tween, and 0.05% NaN_3 in PBS), with agitation. The pins are then immersed in microtiter wells containing 175 µL antisera obtained from human patients diagnosed as having HCV and allowed to incubate at 4°C overnight. The formation of a complex between polyclonal antibodies of the serum and

the polypeptide initiates that the peptides give rise to an immune response in vivo. Such peptides are candidates for the development of vaccines.

Thus, this invention has been described and illustrated. It will be apparent to those skilled in the art that many variations and modifications can be made without departing from the purview of the appended claims and without departing from the teaching and scope of the present invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Tai-An Cha

(ii) TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

(iii) NUMBER OF SEQUENCES: 147

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
(B) STREET: 600 Atlantic Avenue
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02210

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette, 5.25 inch
(B) COMPUTER: IBM compatible
(C) OPERATING SYSTEM: MS-DOS Version 3.3
(D) SOFTWARE: WordPerfect 5.1

TCTCTATCTT CCTCTTGGCT CTGCTGTCC 549

(2) INFORMATION FOR SEQ ID NO: 60

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 549 nucleotides
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: nac5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60

ATGAGCACAA ATCCTAAACC CCAAAGAAAA ACCAAACGTA 40
ACACCAACCG TCGCCCACAG GACGTCAAGT TCCCGGGCGG 80
TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG 120
GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG 160
AGCGGTCGCA ACCTCGTGGA AGGCGACAAC CTATCCCCAA 200
GGCTCGCCGG CCCGAGGGCA GGTCCTGGGC TCAGCCCGGG 240
TACCCTTGGC CCCTCTATGG CAACGAGGGT ATGGGGTGGG 280
CAGGATGGCT CCTGTCACCC CGCGGCTCCC GGCCTAGTTG 320
GGGCCCCACG GACCCCCGGC GTAGGTCGCG TAATTTGGGT 360
AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA 400

TGGGGTACAT TCCGCTCGTC GGCGCCCCCC TAGGGGGCGC 440
TGCCAGGGCC CTGGCACATG GTGTCCGGGT TCTGGAGGAC 480
GGCGTGAACT ATGCAACAGG GAATTTGCCT GGTTGCTCTT 520
TCTCTATCTT CCTCTTGGCT CTGCTGTCC 549

(2) INFORMATION FOR SEQ ID NO: 61

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 549 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: arg2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61

ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA 40
ACACCAACCG CCGCCCACAG GACGTCAAGT TCCCGGGCGG 80
TGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG 120
GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG 160
AGCGGTCGCA ACCTCGTGGA AGGCGACAAC CTATCCCCAA 200
GGCTCGCCAG CCCGAGGGTA GGGCCTGGGC TCAGCCCGGG 240
TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG 280
CAGGGTGGCT CCTGTCCCCC CGCGGCTCCC GGCCTAGTTG 320

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145

CARRAGGAAG AKAGAGAAAG AGCAACCRGG MAR 33

(2) INFORMATION FOR SEQ ID NO: 146

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 nucleotides
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146

AGGCATAGGA CCCGTGTCTT 20

(2) INFORMATION FOR SEQ ID NO: 147

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 nucleotides
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147

CTTCTTTGGA GAAAGTGGTG 20

CLAIMS

1. As a composition of matter, a non-naturally occurring nucleic acid having a non-HCV-1 nucleotide sequence of eight or more nucleotides corresponding to a nucleotide sequence within the hepatitis C virus genome.

2. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome is selected from the regions consisting of the NS5 region, envelope 1 region, 5'UT region, and the core region.

3. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the NS5 region.

4. The composition of claim 3 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome is selected from a sequence within sequences numbered 2-22.

5. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the envelope 1 region.

6. The composition of claim 5 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a sequence within sequence numbers 24-32.

7. The composition of claim 1 wherein at least one sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the 5'UT region.

8. The composition of claim 7 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a sequence within sequences numbered 34-51.

9. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the core region.

10. The composition of claim 9 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a within sequences numbered 53-66.

11. The composition of claim 1 wherein said non-naturally occurring nucleic acid has a nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.

12. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.

13. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, 39-45 in the 5'UT region, and 58-64 in the core region.

14. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.

15. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.

16. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.

17. The composition of claim 1 wherein said non-naturally occurring nucleic acid is capable of priming a reaction for the synthesis of nucleic acid to form a nucleic acid having a nucleotide sequence corresponding to hepatitis C virus.

18. The composition of claim 1 wherein said non-naturally occurring nucleic acid has label means for detecting a hybridization product.

19. The composition of claim 1 wherein said non-naturally occurring nucleic acid has support means for separating a hybridization product from solution.

20. The composition of claim 1 wherein said non-naturally occurring nucleic acid prevents the transcription or translation of viral nucleic acid.

21. A method of forming a hybridization product with a hepatitis C virus nucleic acid comprising the following steps:

a. placing a non-naturally occurring nucleic acid having a nucleotide sequence of eight or more nucleotides corresponding to a non-HCV-1 sequence in the hepatitis C viral genome into conditions in which hybridization conditions can be imposed said non-naturally occurring nucleic acid capable of forming a hybridization product with said hepatitis C virus nucleic acid under hybridization conditions; and

b. imposing hybridization conditions to form a hybridization product in the presence of hepatitis C virus nucleic acid.

22. The method of claim 21 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence in the hepatitis C virus genome corresponds to a sequence within at least one of the regions consisting essentially of NS5 region, envelope 1 region, 5'UT region, and the core region.

23. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within the NS5 region.

24. The method of claim 23 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within sequences numbered 2-22.

25. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within the envelope 1 region.

26. The method of claim 25 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence is selected from a sequence within sequences numbered 24-32.

27. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponding to a sequence within the 5'UT region.

28. The method of claim 27 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence selected from a sequence within sequences numbered 34-51.

29. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponding to a sequence within the core region.

30. The method of claim 29 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence selected from a sequence within sequences numbered 53-66.

31. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.

32. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.

33. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, 39-45 in the 5'UT region, and 58-64 in the core region.

34. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.

35. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.

36. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.

37. The method of claim 21 wherein said hybridization product is capable of priming a reaction for the synthesis of nucleic acid.

38. The method of claim 21 wherein said non-naturally occurring nucleic acid has label means for detecting a hybridization product.

39. The method of claim 21 wherein said non-naturally occurring nucleic acid has support means for separating the hybridization product from solution.

40. The method of claim 21 wherein said non-naturally occurring nucleic acid prevents the transcription or translation of viral nucleic acid.

41. As a composition of matter, a non-naturally occurring polypeptide corresponding to a non-HCV-1 nucleotide sequence of nine or more nucleotides which sequence of nine or more nucleotides corresponds to a sequence within hepatitis C virus genomic sequences.

42. The composition of claim 41 wherein said non-HCV-1 sequence is selected from one of the regions consisting of NS5 region, envelope 1 region, and the core region.

43. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence corresponds to a sequence in the NS5 region.

44. The composition of claim 43 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 2-22.

45. The composition of claim 41 wherein said non-HCV-1 sequence corresponds to a sequence in the envelope 1 region.

46. The composition of claim 45 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 24-32.

47. The composition of claim 41 wherein said non-HCV-1 sequence corresponds to a sequence in the core region.

48. The composition of claim 47 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 52-66.

49. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.

50. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, and 52-57 in the core region.

51. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, and 58-64 in the core region.

52. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, and 65-66 in the core region.

53. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.

54. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.

55. The composition of claim 41 wherein said polypeptide is capable of generating an immune reaction in a host.

56. An antibody capable of selectively binding to the composition of claim 41.

57. A method of detecting one or more genotypes of hepatitis C virus comprising the following steps:

a) placing a non-naturally occurring nucleic acid having a nucleotide sequence of eight or more nucleotides corresponding to one or more genotypes of hepatitis C virus under conditions where hybridization conditions can be imposed,

b) imposing hybridization conditions to form a hybridization product in the presence of hepatitis C virus nucleic acid; and

c) monitoring the non-naturally occurring nucleic acid for the formation of a hybridization product, which hybridization product is indicative of the presence of the genotype of hepatitis C virus.

58. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.

59. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, 39-45 in the 5'UT region, and 58-64 in the core region.

60. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.

61. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.

62. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region.

63. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 67-145.

64. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 69, 71, 73 and 81-99 to identify Group I genotypes in the core and region of the HCV genome.

65. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 70, 72, 70 and 100-118 to identify Group II genotypes in the core and envelope regions of the HCV genome.

66. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 77 to identify Group III genotypes in the 5' UT region of the HCV genome.

67. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence numbered 79 to identify Group IV genotypes in the 5' UT region of the HCV genome.

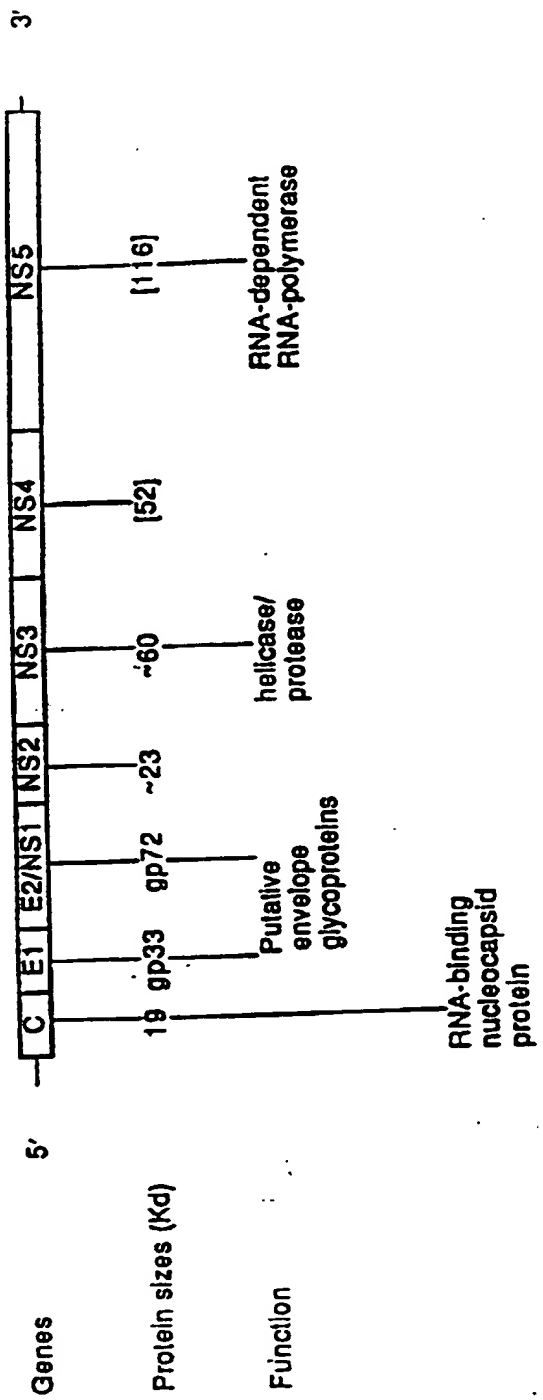
3'
NS5
NS4
NS3
NS2
E2/NS1
E1
C
5'
[116]
[52]
~60
~23
gp72
gp33
19
RNA-dependent RNA-polymerase
helicase/ protease
Putative envelope glycoproteins
RNA-binding nucleocapsid protein
Genes
Protein sizes (Kd)
Function

Fig. 1

Fig. 2a

NS5 REGION

SEQUENCE ID NUMBER	GENOTYPE		
1	GI	1	CTCCACAGTC ACTGAGAGCG ACATCCGTAC GGAGGAGGCA ATCTACCAAT GTTGTGACCT CGACCCCCAA
2		1	CTCCACAGTC ACTGAGAGCG ACATCCGTAC GGAGGAGGCA ATTTACCAAT GTTGTGACCT GGACCCCCAA
3		1	CTCCACAGTC ACTGAGAGCG ACATCCGTAC GGAGGAGGCA ATCTACCAAT GTTGTGATCT GGACCCCCAA
4		1	CTCTACAGTC ACTGAGAACG ACATCCGTAC GGAGGAGGCA ATTTACCAAT GTTGTGACCT GGACCCCCAA
5		1	CTCCACAGTC ACTGAGAGCG ATATCCGTAC GGAGGAGGCA ATCTACCAGT GTTGTGACCT GGACCCCCAA
6		1	CTCTACAGTC ACTGAGAGCG ATATCCGTAC GGAGGAGGCA ATCTACCAAT GTTGTGACCT GGACCCCGAA
7	GII	1	CTCCACAGTC ACTGAGAATG ACACCCGTGT TGAGGAGTCA ATTTACCAAT GTTGTGACTT GGCCCCCGAA
8		1	CTCAACGGTC ACTGAGAATG ACATCCGTGT TGAGGAGTCA ATTTACCAAA GTTGTGACTT GGCCCCCGAG
9		1	CTCAACGGTC ACCGAGAATG ACATCCGTGT TGAGGAGTCA ATTTATCAAT GTTGTGCCTT GGCCCCCGAG
10		1	CTCAACGGTC ACTGAGAGTG ACATCCGTGT CGAGGAGTCG ATTTACCAAT GTTGTGACTT GGCCCCCGAA
11		1	CTCCACAGTC ACTGAGAGTG ACATCCGTGT TGAGGAGTCA ATTTACCAAT GTTGTGACTT GGCCCCCGAA
12		1	CTCAACAGTC ACTGAGAGTG ACATCCGTGT TGAGGAGTCA ATCTACCAAT GTTGTGACTT GGCCCCCGAA
13	GIII	1	CTCAACCGTC ACTGAGAGAG ACATCAGAAC TGAGGAGTCC ATATACCGAG CCTGCTCCCT GCCTGAGGAG
14		1	CTCTACAGTC ACGTAAAAGG ACATCACATC CTAGGAGTCC ATCTACCAGT CCTGTTCACT GCCCGAGGAG
15		1	CTCTACAGTC ACAGAGAGGG ACATCAGAAC CGAGGAGTCC ATCTATCTGT CCTGCTCACT GCCTGAGGAG
16		1	CTCTACAGTC ACGGAGAGGG ACATCAGAAC CGAGGAGTCC ATCTATCTGT CCTGTTCACT GCCTGAGGAG
17		1	CTCAACCGTC ACGGAGAGGG ACATAAGAAC AGAAGAATCC ATATATCAGG GTTGTTCCCT GCCTCAGGAG
18	GV	1	CTCGACCGTT ACCGAACATG ACATAATGAC TGAAGAGTCT ATTTACCAAT CATTGTACTT GCAGCCTGAG
19		1	CTCGACCGTT ACCGAACATG ACATAATGAC TGAAGAGTCC ATTTACCAAT CATTGTACTT GCAGCCTGAG
20	GIV	1	CTCTACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG ATATACCAGT GCTGTAACCT TGAACCGGAG
21		1	CTCGACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG ATATACCAAT GCTGTAACCT TGAACCGGAG
22		1	CTCAACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG ATATACCAAT GCTGTAACCT TGAACCGGAG

SUBSTITUTE SHEET

SUBSTITUTE SHEET

Fig. 2b

NS5 REGION - (2/5)

SEQUENCE ID NUMBER	GENOTYPE		
1	GI	71	GCCCGCGTGG CCATCAAGTC CCTCACCGAG AGGCTTTATG TTGGGGGCCC TCTTACCAAT TCAAGGGGGG
2	GI	71	GCCCGCATGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TCGGGGGCCC TCTTACCAAT TCAAGGGGGG
3	GI	71	GCCCGCGTGG CCATCAAGTC CCTCACTGAG AGGCTTTACG TTGGGGGCCC TCTTACCAAT TCAAGGGGGG
4	GI	71	GCCCGCGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC CCTTACCAAT TCAAGGGGGG
5	GI	71	GCCCGCGTGG CCATCAAGTC CCTCACCGAG AGGCTTTATG TCGGGGGCCC TCTTACCAAT TCAAGGGGGG
6	GI	71	GCCCGTGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC TCTTACCAAT TCAAGGGGGG
7	GII	71	GCCAGACAGG CCATAAGGTC GCTCACAGAG CGGCTCTATG TCGGGGGTCC TATGACTAAC TCCAAAGGGC
8		71	GCCAGACAAG CCATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGCCC CCTGACTAAT TCAAAAGGGC
9		71	GCTAGACAGG CCATAAGGTC GCTCACAGAG CGGCTTTATA TCGGGGGCCC CCTGACCAAT TCAAAGGGGC
10		71	GCCAGGCAGG CCATAAGGTC GCTCACCGAG CGACTTTATA TCGGGGGCCC CCTGACTAAT TCAAAAGGGC
11		71	GCCAGACAGG CTATAAGGTC GCTCACAGAG CGGCTGTACA TCGGGGGTCC CCTGACTAAT TCAAAAGGGC
12		71	GCCAGACAGG CTATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGTCC CCTGACTAAT TCAAAAGGGC
13	GIII	71	GCTCACATTG CCATACACTC GCTGACTGAG AGGCTCTACG TGGGAGGGCC CATGTTCAAC AGCAAGGGCC
14		71	GCTCGAACTG CTATACACTC ACTGACTGAG AGACTATACG TAGGGGGGCC CATGACAAAC AGCAAGGGCC
15		71	GCCCGAACTG CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGGCC CATGACAAAC AGCAAGGGGC
16		71	GCTCGAACTG CCATACACTC ACTGACTGAG AGGCTGTACG TAGGGGGGCC CATGACAAAC AGCAAAGGGC
17		71	GCTAGAACTG CTATCCACTC GCTCACTGAG AGACTCTACG TAGGAGGGCC CATGACAAAC AGCAAGGGAC
18	GV	71	GCGCGTGTGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCCC CATGTATAAC AGCAAGGGGC
19		71	GCACGCGCGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCCC CATGTATAAC AGCAAGGGGC
20	GIV	71	GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAC AGCAAGGGGG
21		71	GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAT AGCAAGGGGG
22		71	GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAC AGCAAGGGGG

Fig. 2c

NS5 REGION - (3/5)

SEQUENCE ID NUMBER	GENOTYPE		
1	GI	141	AGAACTGCGG CTATCGCAGG TGCCGCGCGA GCGGCGTACT GACAACTAGC TGTGGTAACA CCCTCACTTG
2		141	AGAACTGCGG CTACCGCAGG TGCCGCGCGA GCGGCGTACT GACAACTAGC TGTGGTAACA CCCTCACTTG
3		141	AGAACTGCGG CTACCGCAGG TGCCGGGCGA GCGGCGTACT GACAACTAGC TGTGGTAATA CCCTCACTTG
4		141	AAAACTGCGG CTATCGCAGG TGCCGCGCGA GCGGCGTACT GACAACTAGC TGTGGTAACA CCCTCACTTG
5		141	AAAACTGCGG CTATCGCAGG TGCCGCGCAA GCGGCGTACT GACAACTAGC TGTGGTAACA CCCTCACTTG
6		141	AGAACTGCGG CTACCGCAGG TGCCGCGCAA GCGGCGTACT GACGACTAGC TGTGGTAATA CCCTCACTTG
7	GII	141	AGAACTGCGG CTATCGCCGG TGCCGCGCGA GCGGCGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG
8		141	AGAACTGCGG CTATCGCCGA TGCCGCGCCA GCGGTGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG
9		141	AGAACTGCGG TTATCGCCGG TGCCGCGCCA GCGGCGTACT GACGACCAGC TGCGGTAATA CCCTTACATG
10		141	AGAACTGCGG TTATCGCCGG TGCCGCGCGA GCGGCGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG
11		141	AGAACTGCGG CTATCGCCGG TGCCGCGCAA GCGGCGTGCT GACGACTAGC TGCGGTAACA CCCTCACATG
12		141	AGAACTGCGG CTATCGCCGG TGCCGCGCAA GCGGCGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG
13	GIII	141	AGACCTGCGG GTACAGGCGT TGCCGCGCCA GCGGGGTGCT CACCACTAGC ATGGGGAACA CCATCACATG
14		141	AATCCTGCGG GTACAGGCGT TGCCGCGCGA GCGCAGTGCT CACCACCAGC ATGGGCAACA CACTCACGTG
15		141	AATCCTGCGG GTACAGGCGT TGCCGCGCGA GCGGAGTGCT CACCACCAGC ATGGGCAACA CGCTCACGTG
16		141	AATCCTGCGG GTACAGGCGT TGCCGCGCGA GCGGAGTGCT CACCACCAGC ATGGGTAACA CACTCACGTG
17		141	AATCCTGCGG TTACAGGCGT TGCCGCGCCA GCGGGGTCTT CACCACCAGC ATGGGGAATA CCATGACATG
18	GV	141	AACAATGTGG TTATCGTAGA TGCCGCGCCA GCGGCGTCTT CACCACTAGT ATGGGCAACA CCATGACGTG
19		141	AACAATGTGG TTACCGTAGA TGCCGCGCCA GCGGCGTCTT CACCACCAGT ATGGGCAACA CCATGACGTG
20	GIV	141	CCCAGTGTGG TTATCGCCGT TGCCGTGCTA GTGGAGTCCT GCCTACCAGC TTCGGCAACA CAATCACTTG
21		141	CCCAGTGTGG TTATCGCCGT TGCCGTGCTA GTGGAGTTCT GCCTACCAGC TTCGGCAACA CAATCACTTG
22		141	CCCAGTGTGG TTATCGCCGT TGCCGTGCCA GTGGAGTTCT GCCTACCAGC TTCGGCAACA CAATCACTTG

SUBSTITUTE SHEET

Fig. 2d

NS5 REGION - (4/5)

SEQUENCE ID NUMBER	GENOTYPE		
1	GI	211	CTACATCAAG GCCCGGGCAG CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTCGT GTGTGGCGAC
2		211	CTACATCAAG GCCCGGGCAG CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTTGT GTGTGGCGAC
3		211	CTACATCAAG GCCCGGGCAG CCTGTCGAGC CGCAGGGCTC CGGGACTGCA CCATGCTCGT GTGTGGTGAC
4		211	CTACATTAAG GCCCGGGCAG CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTCGT GTGTGGCGAC
5		211	TTACATCAAG GCCCAAGCAG CCTGTCGAGC CGCAGGGCTC CGGGACTGCA CCATGCTCGT GTGTGGCGAC
6		211	TTACATCAAG GCCCGGGCAG CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTCGT GTGTGGCGAC
7	GII	211	CTACCTGAAG GCCACAGCGG CCTGTCGAGC TGCCAAGCTC CAGGACTGCA CGATGCTCGT GAACGGAGAC
8		211	TTACTTGAAG GCCACTGCGG CCTGTAGAGC TGCGAAGCTC CAGGACTGCA CGATGCTCGT GTGCGGAGAC
9		211	TTACTTGAAG GCCTCTGCAG CCTGTCGAGC CGCGAAGCTC CAGGACTGCA CGATGCTCGT GTGTGGGGAC
10		211	TTACTTGAAG GCCTCTGCAG CCTGTCGAGC TGCAAAGCTC CAGGACTGCA CGATGCTCGT GAACGGGGAC
11		211	TTACTTGAAG GCCTCTGCGG CCTGTCGAGC TGCGAAGCTC CAGGACTGCA CGATGCTCGT GTGCGGTGAC
12		211	TTACCTGAAG GCCAGTGCGG CCTGTCGAGC TGCGAAGCTC CAGGACTGCA CAATGCTCGT GTGCGGTGAC
13	GIII	211	CTATGTAAAA GCCCTAGCGG CTTGCAAGGC TGCAGGGATA GTTGCACCCT CAATGCTGGT ATGCGGCGAC
14		211	CTACGTAAAA GCCAGGGCGG CGTGTAACGC CGCGGGGATT GTTGCTCCCA CCATGCTGGT GTGCGGTGAC
15		211	CTACGTGAAA GCCAGAGCGG CGTGTAACGC CGCGGGCATT GTTGCTCCCA CCATGTTGGT GTGCGGCGAC
16		211	CTACGTGAAA GCTAAAGCGG CATGTAACGC CGCGGGCATT GTTGCCCCCA CCATGTTGGT GTGCGGCGAC
17		211	CTACATCAAA GCCCTTGCAG CGTGCAAAGC TGCAGGGATC GTGGACCCTA TCATGCTGGT GTGTGGAGAC
18	GV	211	CTACATTAAG GCTTTAGCCT CCTGTAGAGC CGCAAAGCTC CAGGACTGCA CGCTCCTGGT GTGTGGTGAT
19		211	CTACATCAAG GCTTCAGCCG CCTGTAGAGC TGCAAAGCTC CAGGACTGCA CGCTCCTGGT GTGTGGTGTG
20	GIV	211	TTACATCAAG GCTAGAGCGG CTTCGAAGGC CGCAGGCCTC CGGAACCCGG ACTTTCTTGT CTGCGGAGAT
21		211	TTACATCAAG GCTAGAGCGG CTGCGAAGGC CGCAGGGCTC CGGACCCCGG ACTTTCTCGT CTGCGGAGAT
22		211	TTACATCAAA GCTAGAGCGG CTGCCGAAGC CGCAGGCCTC CGGAACCCGG ACTTTCTTGT CTGCGGAGAT

SUBSTITUTE SHEET

SUBSTITUTE SHEET

Fig. 2e

NS5 REGION - (5/5)

SEQUENCE ID NUMBER	GENOTYPE		
1	GI	281	GACTTAGTCG TTATCTGTGA AAGCGCGGGG GTCCAGGAGG ACGCGGCGAG CCTGAGAGCC
2		281	GACTTAGTCG TTATCTGTGA AAGTGCGGGG GTCCAGGAGG ACGCGGCGAG CCTGAGAGCC
3		281	GACTTGGTCG TTATCTGTGA GAGTGCGGGG GTCCAGGAGG ACGCGGCGAG CCTGAGAGCC
4		281	GACTTAGTCG TTATCTGTGA GAGTGCGGGA GTCCAGGAGG ACGCGGCGAA CTTGAGAGCC
5		281	GACTTAGTCG TTATCTGTGA AAGTCAGGGA GTCCAGGAGG ATGCAGCGAA CCTGAGAGCC
6		281	GACCTAGTCG TTATCTGCGA AAGTGCGGGG GTCCAGGAGG ACGCGGCGAG CCTGAGAGCC
7	GII	281	GACCTTGTCG TTATCTGTGA AAGCGCGGGG AACCAAGAGG ACGCGGCAAG CCTACGAGCC
8		281	GACCTTGTCG TTATCTGTGA AAGCGCGGGA ACCCAGGAGG ATGCGGCGAG CCTACGAGTC
9		281	GACCTTGTCG TTATCTGTGA AAGCGCGGGA ACCCAGGAGG ACGCGGCGAA CCTACGAGTC
10		281	GACCTTGTCG TTATCTGCGA GAGCGCGGGA ACCCAAGAGG ACGCGGCGAG CCTACGAGTC
11		281	GACCTTGTCG TTATCTGTGA GAGCGCGGGA ACCCAAGAGG ACGCGGCGAG CCTACGAGTC
12		281	GACCTTGTCG TTATCTGTGA GAGCGCGGGG ACCCAAGAGG ACGCGGCGAG CCTACGAGTC
13	GIII	281	GACTTAGTTG TCATCTCAGA AAGCCAGGGG ACTGAGGAGG ACGAGCGGAA CCTGAGAGCT
14		281	GACCTGGTCG TCATCTCAGA GAGTCAAGGG GCTGAGGAGG ACGAGCAGAA CCTGAGAGTC
15		281	GACCTGGTTG TCATCTCAGA GAGTCAGGGG GTCGAGGAAG ATGAGCGGAA CCTGAGAGTC
16		281	GACCTAGTCG TCATCTCAGA GAGTCAAGGG GTCGAGGAGG ATGAGCGAAA CCTGAGAGCT
17		281	GACCTGGTCG TCATCTCGGA GAGCGAAGGT AACGAGGAGG ACGAGCGAAA CCTGAGAGCT
18	GV	281	GATCTTGTGG CCATTTGCGA GAGCCAGGGG ACGCACGAGG ATAAAGCGAG CCTGAGAGCC
19		281	ACCTTGGTGG CCATTTGCGA GAGCCAAGGG ACGCACGAGG ATGAAGCGTG CCTGAGAGTC
20	GIV	281	GATCTGGTCG TGGTGGCTGA GAGTGATGGC GTCGACGAGG ATAGAGCAGC CCTGAGAGCC
21		281	GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCGACGAGG ATAGAACAGC CCTGCGAGCC
22		281	GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCAATGAGG ATAGAGCAGC CCTGGGAGCC

340 TOTAL

Fig. 3

ENVELOPE REGION

SEQUENCE ID NUMBER	GENOTYPE		
23	GI	1	GACGGCGTTG GTAATGGCTC AGCTGCTCCG GATCCCACAA GCCATCTTGG ACATGATCGC
24		1	GACGGCGTTG GTGGTAGCTC AGGTACTCCG GATCCCACAA GCCATCATGG ACATGATCGC
25		1	AACGGCGCTG GTAGTAGCTC AGCTGCTCAG GGTCCCGCAA GCCATCGTGG ACATGATCGC
26	GII	1	GACAGCCCTA GTGGTATCGC AGTTACTCCG GATCCCACAA GCCGTCATGG ATATGGTGGC
27		1	AGCAGCCCTA GTGGTGTCGC AGTTACTCCG GATCCCACAA AGCATCGTGG ACATGGTGGC
28		1	GGCAGCCCTA GTGGTGTCGC AGTTACTCCG GATCCCGCAA GCTGTCGTGG ACATGGTGGC
29	GIV	1	TGTGGGTATG GTGGTGGCGC ACGTCCTGCG TTTGCCCCAG ACCTTGTTCG ACATAATAGC
30		1	TGTGGGTATG GTGGTAGCAC ACGTCCTGCG TCTGCCCCAG ACCTTGTTCG ACATAATAGC
31		1	TGTGGGTATG GTGGTGGCGC AAGTCCTGCG TTTGCCCCAG ACCTTGTTCG ACGTGCTAGC
32	GIII	1	TACCACTATG CTCCTGGCAT ACTTGGTGCG CATCCCGGAG GTCATCCTGG ACATTATCAC
23	GI	61	TGGTGCTCAC TGGGGAGTCC TGGCGGGCAT AGCGTATTTC
24		61	TGGAGCCCAC TGGGGAGTCC TGGCGGGCAT AGCGTATTTC
25		61	TGGTGCCCAC TGGGGAGTCC TAGCGGGCAT AGCGTATTTT
26	GII	61	GGGGGCCCAC TGGGGAGTCC TGGCGGGCCT TGCCTACTAT
27		61	GGGGGCCCAC TGGGGAGTCC TGGCGGGCCT TGCTTACTAT
28		61	GGGGGCCCAC TGGGGAATCC TAGCGGGTCT TGCCTACTAT
29	GIV	61	CGGGGCCCAT TGGGGCATCT TGGCGGGCTT GGCCTATTAC
30		61	CGGGGCCCAT TGGGGCATCT TGGCAGGCCT AGCCTATTAC
31		61	CGGGGCCCAT TGGGGCATCT TGGCGGGCCT GGCCTATTAC
32	GIII	61	GGGAGGACAC TGGGGCGTGA TGTTTGGCCT GGCTTATTTC

100 Total

SUBSTITUTE SHEET

WO 92/19743 PCT/US92/04036

11/21

Fig. 4a

5'UT Region

SEQUENCE ID NUMBER	GENOTYPE		
33	GI	1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
34		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
35		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
36		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
37		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
38		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
39	GII	1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
40		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
41		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
42		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
43		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
44		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
45		1	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
46	GIII	1	GCTAGTATCA GTGTCGTACA GCCTCCAGGC CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
47		1	GTTAGTATGA GTCTCGTACA GCCTCCAGGC CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
48	GIV	1	GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CTCCCCCTCC CGGGAGAGCC ATAGTGGTCT
49		1	GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
50	GV	1	GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
51		1	GTTAGTATGA GTGTCGAACA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT

8/21

SUBSTITUTE SHEET

Fig. 4b

5'UT Region (2/5)

SEQUENCE ID NUMBER	GENOTYPE		
33	GI	61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
34		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
35		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
36		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATAAACCC
37		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
38		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATAAACCC
39	GII	61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
40		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
41		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
42		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
43		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
44		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
45		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC
46	GIII	61	GCGGAACCGG TGAGTACACC GGAATTGCCG GGAAGACTGG GTCCTTTCTT GGATAAACCC
47		61	GCGGAACCGG TGAGTACACC GGAATTGCTG GGAAGACTGG GTCCTTTCTT GGATAAACCC
48	GIV	61	GCGGAACCGG TGAGTACACC GGAATCGCTG GGGTGACCGG GTCCTTTCTT GGAGCAACCC
49		61	GCGGAACCGG TGAGTACACC GGAATCGCTG GGGTGACCGG GTCCTTTCTT GGAGTAACCC
50	GV	61	GCGGAACCGG TGAGTACACC GGAATTGCCG GGATGACCGG GTCCTTTCTT GGATAAACCC
51		61	GCGGAACCGG TGAGTACACC GGAATTGCCG GGATGACCGG GTCCTTTCTT GGATAAACCC

SUBSTITUTE SHEET

Fig. 4c

5'UT Region (3/5)

SEQUENCE ID NUMBER	GENOTYPE		
33	GI	121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGTGTTGGGT
34		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGTGTTGGGT
35		121	GCTCAATGCC TGGAGATTTG GGCACGCCCC CGCAAGATCA CTAGCCGAGT AGTGTTGGGT
36		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT
37		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGTGTTGGGT
38		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGTGTTGGGT
39	GII	121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT
40		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT
41		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT
42		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT
43		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT
44		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT
45		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT
46	GIII	121	ACTCTATGCC CGGCCATTTG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGCGTTGGGT
47		121	ACTCTATGCC CAGCCATTTG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGCGTTGGGT
48	GIV	121	GCTCAATACC CAGAAATTTG GGCGTGCCCC CGCGAGATCA CTAGCCGAGT AGTGTTGGGT
49		121	GCTCAATACC CAGAAATTTG GGCGTGCCCC CGCGAGATCA CTAGCCGAGT AGTGTTGGGT
50	GV	121	GCTCAATGCC CGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT
51		121	GCTCAATGCC CGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT

Fig. 4d

ENVELOPE REGION (4/5)

SEQUENCE ID NUMBER	GENOTYPE		
33	GI	181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
34		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
35		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
36		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
37		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
38		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
39	GII	181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
40		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
41		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
42		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
43		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
44		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
45		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
46	GIII	181	TGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
47		181	TGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
48	GIV	181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
49		181	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT

Fig. 4e

5'UT Region (5/5)

SEQUENCE ID NUMBER	GENOTYPE		
33	GI	241	AGACCGTGCA CC
34		241	AGACCGTGCA CC
35		241	AGACCGTGCA CC
36		241	AGACCGTGCA CC
37		241	AGACCGTGCA CC
38		241	AGACCGTGCA CC
39	GII	241	AGACCGTGCA CC
40		241	AGACCGTGCA TC
41		241	AGACCGTGCA CC
42		241	AGACCGTGCA CC
43		241	AGACCGTGCA CC
44		241	AGACCGTGCA CC
45		241	AGACCGTGCA CC
46	GIII	241	AGACCGTGCA TC
47		241	AGACCGTGCA TC
48	GIV	241	AGACCGTGCA AC
49		241	AGACCGTGCA AC

252 Total

SUBSTITUTE SHEET

Fig. 5a

CORE REGION

SEQUENCE ID NUMBER	GENOTYPE		
52	GI	1	ATGAGCACGA ATCCTAAACC TCAAAAAAAA AACAAACGTA ACACCAACCG TCGCCCACAG
53		1	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG
54		1	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG
55		1	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG
56		1	ATGAGCACGA ATCCTAAACC TCAAAGAAGA ACCAAACGTA ACACCAACCG TCGCCCACAG
57		1	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG
58	GII	1	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG
59		1	ATGAGCACAA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG
60		1	ATGAGCACAA ATCCTAAACC CCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG
61		1	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG
62		1	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG
63		1	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG
64		1	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG
65	GIII	1	ATGAGCACAA ATCCTAAACC TCAAAGAAAA ACCAAAAGAA ACACTAACCG CCGCCCACAG
66		1	ATGAGCACAA ATCGTCAACC TCAAAGAAAA ACCAAAAGAA ACACTAACCG CCGCCCACAG

SUBSTITUTE SHEET

Fig. 5b

CORE REGION (2/9)

SEQUENCE ID NUMBER	GENOTYPE		
52	GI	61	GACGTCAAGT TCCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG
53		61	GACGTCAAGT TCCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG
54		61	GACGTTAAGT TCCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG
55		61	GACGTCAAGT TCCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG
56		61	GACGTCAAGT TCCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG
57		61	GACGTCAAGT TCCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG
58	GII	61	GACGTTAAGT TCCCGGGCGG TGGCCAGGTC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG
59		61	GACGTCAAGT TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG
60		61	GACGTCAAGT TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG
61		61	GACGTCAAGT TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG
62		61	GACGTCAAGT TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG
63		61	GACGTCAAGT TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG
64		61	GACGTCAAGT TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG
65	GIII	61	GACGTCAAGT TCCCGGGCGG TGGCCAGATC GTTGGCGGAG TATACTTGCT GCCGCGCAGG
66		61	GACGTCAAGT TCCCGGGCGG TGGTCAGATC GTTGGCGGAG TATACTTGTT GCCGCGCAGG

SUBSTITUTE SHEET

Fig. 5c

CORE REGION (3/9)

SEQUENCE ID NUMBER	GENOTYPE		
52	GI	121	GGCCCTAGAT TGGGTGTGCG CGCGACGAGA AAGACTTCCG AGCGGTCGCA ACCTCGAGGT
53		121	GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT
54		121	GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT
55		121	GGCCCTAGAT TGGGTGTGCG CACGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT
56		121	GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT
57		121	GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGT
58	GII	121	GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA
59		121	GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA
60		121	GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA
61		121	GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA
62		121	GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA
63		121	GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA
64		121	GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA
65	GIII	121	GGCCCGAGAT TGGGTGTGCG CGCGACGAGG AAAACTTCCG AACGATCCCA GCCACGCGGA
66		121	GGCCCCAGGT TGGGTGTGCG CGCGACGAGG AAAACTTCCG AACGGTCCCA GCCACGTGGG

SUBSTITUTE SHEET

Fig. 5d

CORE REGION (4/9)

SEQUENCE ID NUMBER	GENOTYPE		
52	GI	181	AGACGTCAGC CTATCCCCAA GGCTCGTCGG CCCGAGGGCA GGACCTGGGC TCAGCCCGGG
53		181	AGACGTCAGC CTATCCCCAA GGCGCGTCGG CCCGAGGGCA GGACCTGGGC TCAGCCCGGG
54		181	AGACGTCAGC CTATCCCTAA GGCGCGTCGG CCCGAGGGCA GGACCTGGGC TCAGCCCGGG
55		181	AGACGTCAGC CCATCCCCAA GGCTCGTCGA CCCGAGGGCA GGACCTGGGC TCAGCCCGGG
56		181	AGACGTCAGC CTATCCCCAA GGCACGTCGG CCCGAGGGTA GGACCTGGGC TCAGCCCGGG
57		181	AGACGCCAGC CTATCCCCAA GGCGCGTCGG CCCGAGGGCA GGACCTGGGC TCAGCCCGGG
58	GII	181	AGGCGACAAC CTATCCCCAA GGCTCGCCAG CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG
59		181	AGGCGACAAC CTATCCCCAA GGCTCGCCAG CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG
60		181	AGGCGACAAC CTATCCCCAA GGCTCGCCGG CCCGAGGGCA GGTCCTGGGC TCAGCCCGGG
61		181	AGGCGACAAC CTATCCCCAA GGCTCGCCAG CCCGAGGGTA GGGCCTGGGC TCAGCCCGGG
62		181	AGGCGACAAC CTATCCCCAA GGCTCGCCGG CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG
63		181	AGGCGACAAC CTATCCCCAA GGCTCGCCGG CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG
64		181	AGGCGACAAC CTATCCCCAA GGCTCGCCAG CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG
65	GIII	181	AGGCGTCAGC CCATCCCTAA AGATCGTCGC ACCGCTGGCA AGTCCTGGGG AAGGCCAGGA
66		181	AGGCGCCAGC CCATCCCCAA AGATCGGCGC ACCACTGGCA AGTCCTGGGG GAAGCCAGGA

SUBSTITUTE SHEET

Fig. 5e

CORE REGION (5/9)

SEQUENCE ID NUMBER	GENOTYPE		
52	GI	241	TACCCTTGGC CCCTCTATGG CAATGAGGGC TGCGGGTGGG CGGGATGGCT CCTGTCTCCC
53		241	TACCCTTGGC CCCTCTATGG CAATGAGGGT TGCGGGTGGG CGGGATGGCT CCTGTCTCCC
54		241	TACCCCTGGC CCCTCTATGG TAATGAGGGT TGCGGATGGG CGGGATGGCT CCTGTCCCCC
55		241	TACCCTTGGC CCCTCTATGG CAATGAGGGC TGCGGGTGGG CGGGATGGCT CCTGTCTCCC
56		241	TACCCTTGGC CCCTCTATGG CAATGAGGGT TGCGGGTGGG CGGGATGGCT CCTGTCTCCC
57		241	TACCCTTGGC CCCTCTATGG CAATGAGGGT TGCGGGTGGG CGGGATGGCT CCTGTCTCCC
58	GII	241	TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGATGGCT CCTGTCACCC
59		241	TACCCTTGGC CCCTCTATGG CAACGAGGGT ATGGGGTGGG CAGGATGGCT CCTGTCACCC
60		241	TACCCTTGGC CCCTCTATGG CAACGAGGGT ATGGGGTGGG CAGGATGGCT CCTGTCACCC
61		241	TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGGTGGCT CCTGTCCCCC
62		241	TATCCTTGGC CCCTCTATGG CAATGAGGGT CTGGGGTGGG CAGGATGGCT CCTGTCACCC
63		241	TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGATGGCT CCTGTCACCC
64		241	TACCCCTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGATGGCT CCTGTCACCC
65	GIII	241	TATCCTTGGC CCCTGTATGG GAATGAGGGT CTCGGCTGGG CAGGGTGGCT CCTGTCCCCC
66		241	TACCCTTGGC CCCTGTATGG GAATGAGGGT CTCGGCTGGG CAGGGTGGCT CCTGTCCCCC

SUBSTITUTE SHEET

Fig. 5f

CORE REGION (6/9)

SEQUENCE ID NUMBER	GENOTYPE		
52	GI	301	CGTGGCTCTC GGCCTAGCTG GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT
53		301	CGTGGCTCTC GGCCTAGTTG GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT
54		301	CGTGGCTCTC GGCCTAGTTG GGGCCCTACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT
55		301	CGTGGCTCTC GGCCTAGCTG GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT
56		301	CGCGGCTCTC GGCCTAACTG GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT
57		301	CGTGGCTCTC GGCCTAGCTG GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT
58	GII	301	CGTGGCTCTC GGCCTAGTTG GGGCCCCACG GACCCCCGGC GTAGGTCGCG TAATTTGGGT
59		301	CGTGGCTCTC GGCCTAGTTG GGGCCCCACG GACCCCCGGC GTAGGTCGCG TAATTTGGGT
60		301	CGCGGCTCCC GGCCTAGTTG GGGCCCCACG GACCCCCGGC GTAGGTCGCG TAATTTGGGT
61		301	CGCGGCTCCC GGCCTAGTTG GGGCCCCACA GACCCCCGGC GTAGGTCGCG TAATTTGGGT
62		301	CGCGGCTCTC GGCCTAGCTG GGGCCCTACC GACCCCCGGC GTAGGTCGCG CAACTTGGGT
63		301	CGTGGTTCTC GGCCTAGTTG GGGCCCCACG GACCCCCGGC GTAGGTCGCG CAATTTGGGT
64		301	CGCGGCTCCC GGCCTAGTTG GGGCCCCAAA GACCCCCGGC GTAGGTCGCG TAATTTGGGT
65	GIII	301	CGTGGCTCTC GCCCTTCATG GGGCCCCACT GACCCCCGGC ATAGATCGCG CAACTTGGGT
66		301	CGCGGTTCTC GCCCTTCATG GGGCCCCACT GACCCCCGGC ATAGATCACG CAACTTGGGT

Fig. 5g

CORE REGION (7/9)

SEQUENCE ID NUMBER	GENOTYPE		
52	GI	361	AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT ACCGCTCGTC
53		361	AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCACA TGGGGTACAT ACCGCTCGTC
54		361	AAGGTCATCG ATACCCTCAC GTGCGGCTTC GCCGACCACA TGGGGTACAT TCCGCTCGTT
55		361	AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT ACCGCTCGTC
56		361	AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT ACCGCTCGTC
57		361	AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT ACCGCTCGTC
58	GII	361	AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTCGTC
59		361	AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTTGTC
60		361	AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTCGTC
61		361	AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTCGTC
62		361	AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTCGTC
63		361	AAGATCATCG ATACCCTCAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTCGTC
64		361	AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTCGTC
65	GIII	361	AAGGTCATCG ATACCCTAAC GTGCGGTTTT GCCGACCTCA TGGGGTACAT TCCCGTCATC
66		361	AAGGTCATCG ATACCCTAAC GTGTGGTTTT GCCGACCTCA TGGGGTACAT TCCCGTCGGT

Fig. 5h

CORE REGION (8/9)

SEQUENCE ID NUMBER	GENOTYPE		
52	GI	421	GGCGCCCCTC TTGGAGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC
53		421	GGCGCCCCTC TTGGAGGCGC TGCCAGGGCT CTGGCGCATG GCGTCCGGGT TCTGGAAGAC
54		421	GGCGCCCCTC TTGGGGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC
55		421	GGCGCCCCTC TTGGAGGCGC TGCCAGAGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC
56		421	GGCGCCCCTC TTGGAGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC
57		421	GGCGCCCCTC TTGGAGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC
58	GII	421	GGCGCCCCCC TTAGGGGCGC TGCCAGGGCC TTGGCGCATG GCGTCCGGGT TCTGGAGGAC
59		421	GGCGCCCCCC TAGGGGGCGC TGCCAGGGCC CTGGCACATG GTGTCCGGGT TCTGGAGGAC
60		421	GGCGCCCCCC TAGGGGGCGC TGCCAGGGCC CTGGCACATG GTGTCCGGGT TCTGGAGGAC
61		421	GGCGCCCCCC TAGGGGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGAC
62		421	GGCGCCCCCC TTAGGGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGAC
63		421	GGCGCCCCCC TAGGGGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGAC
64		421	GGCGCCCCCT TAGGGGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGAC
65	GIII	421	GGCGCCCCCG TTGGAGGCGT TGCCAGAGCT CTCGCCCACG GAGTGAGGGT TCTGGAGGAT
66		421	GGTGCCCCCG TTGGTGGTGT CGCCAGAGCC CTTGCCCATG GGGTGAGGGT TCTGGAAGAC

12/102

SUBSTITUTE SHEET

SUBSTITUTE SHEET

Fig. 5i

CORE REGION (9/9)

SEQUENCE ID NUMBER	GENOTYPE		
52	GI	481	GGCGTGAACT ATGCAACAGG GAACCTTCCT GGTTGCTCTT TCTCTATCTT CCTTCTGGCC CTGCTCTCT
53		481	GGCGTGAACT ATGCAACAGG GAACCTTCCT GGTTGCTCTT TCTCTATCTT CCTTCTGGCC CTGCTCTCT
54		481	GGCGTGAACT ATGCAACAGG GAATCTTCCT GGTTGCTCTT TCTCTATCTT CCTTCTGGCC CTTCTCTCT
55		481	GGCGTGAACT ATGCAACAGG GAACCTTCCC GGTTGCTCTT TCTCTATCTT CCTTCTGGCC CTGCTCTCT
56		481	GGCGTGAACT ATGCAACAGG GAACCTTCCT GGTTGCTCTT TCTCTATCTT CCTTCTGGCC CTGCTCTCT
57		481	GGCGTGAACT ATGCAACAGG GAACCTTCCT GGTTGCTCTT TTTCTATTTT CCTTCTGGCC CTGCTCTCT
58	GII	481	GGCGTGAACT ACGCAACAGG GAATCTGCCC GGTTGCTCCT TTTCTATCTT CCTCTTGGCT CTGCTGTCC
59		481	GGCGTGAACT ATGCAACAGG GAATTTGCCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT CTGCTGTCC
60		481	GGCGTGAACT ATGCAACAGG GAATTTGCCT GGTTGCTCTT TCTCTATCTT CCTCTTGGCT CTGCTGTCC
61		481	GGCGTGAACT ATGCAACAGG GAATCTGCCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT TTGCTGTCC
62		481	GGCGTGAACT ATGCAACAGG GAATTTGCCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT TTGCTGTCC
63		481	GGCGTGAACT ATGCAACAGG GAATCTGCCC GGTTGCTCCT TTTCTATCTT CCTTCTGGCT TTGCTGTCC
64		481	GGCGTGAACT ATGCAACAGG GAATCTACCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT TTGCTGTCC
65	GIII	481	GGGGTAAATT ATGCAACAGG GAATTTGCCC GGTTGCTCTT TCTCTATCTT TCTCTTAGCC CTCTTGTCT
66		481	GGGATAAATT ATGCAACAGG GAATCTGCCC

549 Total